

MaRDI Workshop on Scientific Computing

Wednesday 26 October 2022 - Friday 28 October 2022

Münster Mathematics Conference Centre

Book of Abstracts

Contents

Dark Data in Mathematics <i>Björn Schembera, Martin Thomas Horsch</i>	1
Helping Ontology Extension with Natural Language Processing for Catalysis <i>Alexander Behr, Marc Völkenrath, Norbert Kockmann</i>	1
Managing reproducibility in computational experiments <i>Emil Løvbak, Mustafa Dikmen, Naeem Muhammad, Giovanni Samaey</i>	2
Expanding an Ontology with semantically linked CFD-Simulation Data by Segmentation into reusable Concepts <i>Alexander Behr, Norbert Kockmann</i>	2
Reproducibility as a service: collaborative scientific computing with Julia <i>Michael Schlottke-Lakemper, Hendrik Ranocha</i>	3
Increasing the reproducibility of scientific results in mathematics and related fields: Expe- riences and discussions with the research community <i>Christian Riedel, Ulrike Lucke, Ralf Engbert</i>	4
Reproducibility Infrastructure of the Julia Language <i>Jürgen Fuhrmann</i>	4
Developing a Sustainable and FAIR HPC Sparse Linear Algebra Framework <i>Terry Cojean</i>	5
The Importance of Symbolic Data Types <i>Antony Della Vecchia</i>	5
Confirmable Workflows in Computer Algebra <i>Lars Kastner Antony Della Vecchia</i>	5
Benchmarking supervised regression algorithms with mlr3 and OpenML <i>Sebastian Fischer, Michel Lang</i>	6
Experiences in Refactoring Software Code for the Solution of the Wave Equation based on a Very Weak Space-Time Variational Formulation <i>Jürgen Vorloeper</i>	6
Challenges and tools for FAIR data in the heterogeneous CRC 1456 ”Mathematics of Exper- iment” <i>Christoph Lehrenfeld, Markus Osterhoff</i>	6

xSDK: an Ecosystem of Interoperable Independently Developed Math Libraries <i>Ulrike Meier Yang</i>	7
Fostering interdisciplinary research by composable Julia software <i>Michael Herbst</i>	7
Graph-based Data Representation for Crash-worthiness Simulations <i>Anahita Pakiman, Jochen Garcke</i>	8
Registration and Welcome	9
Reception	9
Presentation of Workgroups	9
Conference Dinner	9
Lunch & End of Workshop	10
Creating sustainable research software by the example of the deal.II library <i>Martin Kronbichler</i>	10
FitBenchmarking: an open source tool for comparing data analysis software <i>Tyrone Rees, Andrew Lister, Anders Markvardsen, Jaroslav Fowkes, Tim Snow</i>	10
Building a Knowledge Graph for Scientific Computing <i>Frank Wübbeling, René Fritze</i>	11
Design of a Workflow Description for Documentation and Integration of FAIR Computational Experiments <i>Pavan Veluvalli, Jan Heiland, Peter Benner</i>	11
ADOL-C: 40 years of software development <i>Andrea Walther</i>	12
Towards a Benchmark Framework for Model Order Reduction in the Mathematical Research Data Initiative (MaRDI) <i>Peter Benner, Christian Himpe, Kathryn Lund, Jens Saak</i>	13
MaRDI - The Mathematical Research Data Initiative within the German National Research Data Infrastructure (NFDI) <i>Thomas Koprucki</i>	13
Metadata4Ing: An ontology for describing the generation and provenance of research data within a scientific activity <i>Dorothea Iglezakis, Benjamin Farnbacher</i>	14
preCICE – A General-Purpose Simulation Coupling Interface <i>Benjamin Uekermann</i>	14
Towards Foundations for Open Interfaces for Scientific Computing <i>René Fritze, Mario Ohlberger, Stephan Rave</i>	15

Contributed Talks / 3**Dark Data in Mathematics****Authors:** Björn Schembera¹; Martin Thomas Horsch²¹ IANS / University of Stuttgart² Department of Data Science, Norwegian University of Life Sciences**Corresponding Author:** bjoern.schembera@mathematik.uni-stuttgart.de

Dark data is data that is poorly managed [1, 2]. It is diametrically opposed to FAIR data because its epistemic status is unclear, and it is neither findable, accessible, interoperable, nor reusable. For example, research data may be uncurated, unavailable, unannotated, biased, or incomplete. Examples of dark data in scientific computing include the vast amounts of data that are held unavailable, unsearchable, and unannotated on the parallel file systems or tape archive storages of high-performance computing centers [2]. But what is dark data in mathematics, i.e. dark data concerning mathematical research data such as models, formulas and other abstract artefacts [3]? Can these also be unavailable, unsearchable or unannotated, depending on their specifics? The talk will explore the extent to which even these more abstract types of assets and research data in mathematics can become dark in the absence of good research data management practice.

Dark data as a negative category serves as an analytical tool to work out how the data management processes can be improved and aligned with the FAIR principles. It is a requirement for research data infrastructures that they enable good practices in research data management [4]. For this purpose, we propose a strategy based on metadata standardization by ontologies expressed simultaneously and consistently in OWL description logic (to permit SPARQL queries, etc.) and first order logic (e.g., to facilitate answer set programming) [5]. Our primary aim is to support the documentation of epistemic metadata [6], i.e., information about the knowledge status of data, which we propose to standardize by a mid-level ontology [7]. Below this, at the domain level, documenting the epistemic metadata in a way that is adequate for each academic community will require an evaluation of disciplinary scientific practices and conventions in detail [6]. It is here that NFDI consortia can provide dedicated support by encouraging academic communities to reflect upon their own practices and engage in community-driven metadata standardization. The key objective is for all research data to attain epistemic FAIRness: A status where it is accessible and intelligible to all stakeholders in what way the data has been given an interpretation as knowledge, making that knowledge and its reuse machine-actionable and thereby avoiding that it falls into darkness.

[1] Heidorn, P. B. "Shedding light on the dark data in the long tail of science." *Library Trends* 57.2 (2008): 280-299.

[2] Schembera, B., and Durán, J. M. "Dark data as the new challenge for big data science and the introduction of the scientific data officer." *Philosophy & Technology* 33.1 (2020): 93-115.

[3] Koprucki, T., Tabelow, K., and Kleinod, I. "Mathematical research data." *PAMM* 16.1 (2016): 959-960.

[4] Horsch, M. T., et al. "Interoperability and architecture requirements analysis and metadata standardization for a research data infrastructure in catalysis." In Pozanenko, A., et al. (eds.), *Proceedings of DAMDID/RCDL 2021*, CCIS no. 1620, Springer, 2022.

[5] Horsch, M. T. "Mereosemiotics: Parts and signs." In Sanfilippo, E. M., et al. (eds.), *Proceedings of JOWO 2021*, CEUR no. 2969, CEUR-WS, 2021.

[6] Schembera, B., and Horsch, M. T. "Dark data and epistemic metadata in molecular modelling." Submitted, 2022.

[7] Horsch, M. T., and Schembera, B. "Documentation of epistemic metadata by a mid-level ontology of cognitive processes." Submitted (preprint doi:10.5281/zenodo.6638457), 2022.

Contributed Talks / 4**Helping Ontology Extension with Natural Language Processing for Catalysis**

Author: Alexander Behr¹

Co-authors: Marc Völkenrath²; Norbert Kockmann³

¹ TUDO-NFDI4Cat

² Tu Dortmund

³ TU Dortmund University

Corresponding Author: alexander.behr@tu-dortmund.de

Ontologies store semantic knowledge in a machine-readable way and represent domain knowledge in controlled vocabulary. Scientific results often are published in text form, thus discouraging research data FAIRness. Using natural language processing (NLP), concept names and relations can be extracted from text datasets.

A workflow to process scientific textual text corpora is introduced regarding catalysis research. NLP techniques are used to vectorize the textual data. This allows for hierarchical clustering of concepts, also yielding concept names. In addition, ontologies containing the resulting concept names are searched from a database. Once found, corresponding existing definitions of those concepts are also important output enabling domain experts to validate correctly found ontology classes. Subsequently performed hierarchical clustering of the concept names based on the text corpora prepares the found data for ontology matching, assisting in ontology extension. Previously undefined concepts and unstructured relations can thus be more easily introduced into existing ontologies based on their descriptive scientific texts. A structured extension of ontologies supported by NLP methods is thus made possible to facilitate FAIR data management workflow. The contribution shows successful applications and highlights existing hurdles, too.

Contributed Talks / 5

Managing reproducibility in computational experiments

Authors: Emil Løvbak¹; Mustafa Dikmen¹; Naeem Muhammad¹; Giovanni Samaey¹

¹ KU Leuven

Corresponding Author: emil.loevbak@kuleuven.be

It is often difficult to reproduce computational experiments from papers due to a lack of detailed in how such experiments are documented. Even when researchers publish their code along side a paper, key information is often not well documented: *What version of an external software library was used? What value should be given to an undocumented model parameter? Which specific version of the code produced the results?*

In the NUMA research group at KU Leuven, we are developing a set of workflows for software development and computational experiments which attach this key information to our computational results as metadata. On a technical front, we have integrated Git with the iRODS data management software. We will reproduce our experimental results by reconstructing the exact computational environment that we used to produce our results. We will employ a Docker based reproducibility package for this purpose. In this presentation we will present our technical solutions, as well as our experiences in integrating these technical approaches with researchers' workflows.

Contributed Talks / 6

Expanding an Ontology with semantically linked CFD-Simulation Data by Segmentation into reusable Concepts

Author: Hendrik Borgelt^{None}

Co-authors: Alexander Behr ¹; Norbert Kockmann ²

¹ *TUDO-NFDI4Cat*

² *TU Dortmund University*

Corresponding Author: hendrik.borgelt@tu-dortmund.de

Computational-Fluid-Dynamics (CFD) simulation and other numerical simulation tools generate a rich variety of complex (meta-)data, which are inherently difficult to store in a FAIR manner (Findable-Accessible-Interoperable-Reusable). As the amount of data generated by such simulations is one of the major challenges, meta-data, e.g. the simulation settings and major output variables, offers the possibility of restoring, revising, and re-evaluating existing simulations. However, due to the linked nature of the parameters of such simulations, classifying and storing metadata in a standardized manner is difficult. Ontologies are key to the FAIRness of such data, as they inherently classify the data and are capable of reasoning and querying.

As storing large amounts of linked data in ontologies comes with its challenges, a segmentation method is introduced for data condensation and pre-classification. The method proposed here uses nested python dictionaries in the form of JSON files and populates an existing ontology with respective simulation data. Those nested dictionaries represent the linked structure of the setting options and are either given or can be generated from existing simulations (e.g. the CFX command language from ANSYS). These dictionaries are segmented into sub-dictionaries, representing main concepts, which are then archived and related between different simulation dictionaries to pre-classify the data. The population of the ontology is performed via the above-mentioned sub-dictionaries. This condenses the data by linking and reusing concepts between multiple simulations.

While this method is generic in its concept, the workflow has already been performed by converting the results of simulations into meta-data, populating an ontology with such data, and evaluating the results. Important steps in the workflow are already solved, such as the population of arbitrarily named entities that occur throughout the dictionaries, the multiplicity of concepts, varying linkages, and renaming into semantically aligned classes.

As the number of manual inputs is minimized, which are required to populate the ontology with the given data, a non-expert operable and FAIRer storage was achieved.

Contributed Talks / 8

Reproducibility as a service: collaborative scientific computing with Julia

Authors: Michael Schlottke-Lakemper¹; Hendrik Ranocha²

¹ *RWTH Aachen University*

² *University of Hamburg*

Corresponding Authors: mail@ranocha.de, m.schlottke-lakemper@hirs.de

With the complexity of the involved algorithms and software packages, reproducibility of numerical simulations is often difficult to achieve. This makes it harder to collaborate on research projects, since there can be a considerable ramp-up time for new project members before they are able to contribute to a joint code base. Julia is a modern, dynamic programming language designed for high-performance scientific computing. It makes it easy to set up collaborative development workflows by providing tools to create fully reproducible environments for all major operating systems, which can then be easily shared in a small Git repository. In this talk, we will give a brief general introduction to Julia and its capabilities, focusing on those aspects that make it interesting for collaborative research software development. We will include real-world examples from our research code Trixi.jl, a Julia package for adaptive numerical simulations of fluid flow and other conservation laws.

Contributed Talks / 9

Increasing the reproducibility of scientific results in mathematics and related fields: Experiences and discussions with the research community

Authors: Christian Riedel¹; Ulrike Lucke¹; Ralf Engbert¹

¹ *University of Potsdam*

Corresponding Author: christian.riedel@uni-potsdam.de

Reproducible research results are vital to safeguard scientific quality assurance and to build a reliable foundation for sustainable research. The discussions on this issue accelerated when investigations on reproducibility showed that few scientific publications across many research fields allow for reproducing the published results. This reproducibility crisis is well known within the respective communities. However, through various procedures, such as introducing policies by scientific journals and funding agencies, establishing institutional support structures, and forming national initiatives, various stakeholders in science make efforts to engage on this matter.

Within this context, the Collaborative Research Center (CRC) 1294 – Data Assimilation applies state-of-the-art measures to support associated scientists in their research data management. The measures generally address the reproducibility of published research results and encompass the provision of an IT infrastructure for collaborative work and workshops for knowledge perpetuation. When we investigated the reproducibility of 108 papers published between 2017 and 2021 by the CRC’s researchers, we found that the reproducibility rate increased over time. We associate our support structures in research data management and certain changes in research culture with this improvement. However, many publications did not allow for reproducing the published results, and the overall reproducibility rate and reasons for failed reproducibility correspond to previous investigations on reproducible science.

Since the CRC is located in applied mathematics and related fields, this naturally addresses the research culture in this area. Based on our experiences, we conclude that mandatory artifact sharing, support structures for scientists, the improvement of data quality, and the recognition of research data as scientific achievements are vital elements in improving the reproducibility of scientific results. Furthermore, we strongly recommend that researchers proposing new algorithms support their theoretical publications with computer code. Since our conclusions involve the practices of researchers in mathematics and related fields, we aim to bring the discussion to the community. We present our findings and examine the viewpoints of researchers to incorporate the community’s interests in future measures to improve the reproducibility of research results.

Contributed Talks / 10

Reproducibility Infrastructure of the Julia Language

Author: Jürgen Fuhrmann^{None}

Corresponding Author: juergen.fuhrmann@wias-berlin.de

The Julia language is mostly advertised for the underlying vision to provide an environment for scientific computing and data science which allows to implement algorithms using a syntax similar to Python and Matlab but without sacrificing performance.

Reproducibility and reusability are further important aspects of Julia and its ecosystem.

Julia’s built-in package manager Pkg.jl provides tools to exactly reproduce project environments. Semantic Versioning and maintenance of compatibility constraints are mandatory for packages available from the Julia General registry. Automatic package management is built as well into Julia’s Pluto.jl computational notebooks. Julia’s BinaryBuilder allows to maintain binary packages for all relevant platforms supported by Julia itself. An artifact handling system handles access to artifacts stored outside the Julia ecosystem and their versioning.

The talk will start with highlighting the advantages of avoiding the two-language problem – another vision behind Julia – under the aspect of reproducibility. It will give a pragmatic overview on the topics mentioned from the perspective of a Julia user and co-developer of a group of Julia packages for the numerical solution of partial differential equations.

The talk will start with highlighting the advantages of avoiding the two-language problem – another vision behind Julia – under the aspect of reproducibility. It will give an pragmatic overview on the topics mentioned from the perspective of a Julia user and co-developer of a group of Julia packages for the numerical solution of partial differential equations.

Invited Talks / 11

Developing a Sustainable and FAIR HPC Sparse Linear Algebra Framework

Authors: Terry Cojean¹;

¹ *Karlsruhe Institute of Technology*

Corresponding Author: terry.cojean@kit.edu

With a strong reliance on research software projects in both industry and for scientific simulations, research software sustainability is increasingly becoming a major point of contention. A necessary but nonsufficient aspect of software sustainability is Continuous Integration and Benchmarking (CI/CB/Cx). In addition, software flexibility to support newer HPC hardware as well as modern, flexible interface are also necessary for sustainability. Finally, a mathematical and HPC software's testing strategies can be complex due to the different hardware behavior, and the need to ensure numerical accuracy. In this talk, we will showcase the design of the Ginkgo sparse linear algebra framework which features good software design techniques, was designed with testability, benchmarking as well as Cx practices as centerpieces.

Contributed Talks / 12

The Importance of Symbolic Data Types

Author: Antony Della Vecchia^{None}

Corresponding Author: vecchia@math.tu-berlin.de

Convex hull computations are an essential part of many scientific calculations. We present an experiment written in Julia involving convex hull computations done with two different types of data, floats and rationals. A comparison of the results shows that using floats leads to the loss of the combinatorics of the experiment.

Contributed Talks / 13

Confirmable Workflows in Computer Algebra

Authors: Lars Kastner¹; Antony Della Vecchia^{None}

¹ *TU Berlin*

Corresponding Author: kastner@math.tu-berlin.de

Computer experiments are becoming an essential part of pure math fields, such as combinatorics, commutative algebra and algebraic geometry. We discuss the arising challenges and the work of the task area on computer algebra of MaRDI.

Contributed Talks / 14

Benchmarking supervised regression algorithms with mlr3 and OpenML

Authors: Sebastian Fischer¹; Michel Lang²

¹ *LMU Munich*

² *TU Dortmund*

Corresponding Author: sebf.fischer@gmail.com

Machine learning research should be easily accessible and reusable. OpenML is an open platform for sharing datasets, algorithms, and experiments. mlr3 is an open-source collection of R packages providing a unified interface for machine learning in the R language. One of the projects in the MaRDI task area 3 (statistics and machine learning) was the interface package mlr3oml which allows for seamless integration between these two components. This presentation will show the work on this project and demonstrate an example workflow of benchmarking supervised regression algorithms.

Contributed Talks / 15

Experiences in Refactoring Software Code for the Solution of the Wave Equation based on a Very Weak Space-Time Variational Formulation

Author: Jürgen Vorloeper^{None}

Corresponding Author: juergen.vorloeper@hs-ruhrwest.de

Recently a mathematical approach for the efficient numerical solution of the Wave Equation based on a very weak space-time variational formulation has been proposed by J. Henning, D. Palitta, V. Simoncini and K. Urban. Beside mathematical analysis the authors developed software code generating numerical results. This software code is actually in progress of refactoring to facilitate further developments with respect to mathematical methodologies and usage in combination with different software packages. Special care needs the design and implementation of efficient preconditioning strategies for the algebraic linear equations. The solution of (generalized) Sylvester Equations plays an important role in this context.

In this talk we address experiences in refactoring software code including mathematical challenges, software design and reproduction of numerical results.

Invited Talks / 16

Challenges and tools for FAIR data in the heterogeneous CRC 1456 "Mathematics of Experiment"

Author: Christoph Lehrenfeld¹

Co-authors: Pedro Klein ¹; Markus Osterhoff ¹

¹ *University of Göttingen*

Corresponding Author: lehrenfeld@math.uni-goettingen.de

Within the collaborative research center “CRC 1456 - Mathematics of Experiment” of the German Research Foundation (DFG) several research groups from the natural sciences and mathematics jointly work on measurement and extracting the most information from them. These measurement data come from different types of measurements ranging from nanoscale imaging to observations of the Sun. As the different type of data sets, the involved methods and algorithms to retrieve information from the corresponding measurement data is diverse. Nevertheless, the CRC is committed to meet the highest standards in reproducibility and accessibility. To this end we apply and develop solutions that work for the whole CRC and identify subgroups with similarities in data types or algorithms. In this talk we present the challenges that come with the involvement of many different scientific communities and diverse data types and software and discuss different measures that are undertaken within the CRC

- to ensure FAIR research data handling and
- to increase accessibility of research methods and results.

Invited Talks / 17

xSDK: an Ecosystem of Interoperable Independently Developed Math Libraries

Author: Ulrike Meier Yang¹

¹ *Lawrence Livermore National Laboratory*

Corresponding Author: yang11@llnl.gov

The development of emerging extreme-scale architectures with higher performance potential provides developers of application codes, including multiphysics modeling, and the coupling of simulations and data analytics, unprecedented resources for larger simulations achieving more accurate solutions than ever before. Achieving high performance on these new heterogeneous architectures requires expertise knowledge. To meet these challenges in a timely fashion and make the best use of these capabilities requires a variety of mathematical libraries that are developed by diverse independent teams throughout the HPC community. It is not sufficient for these libraries to individually deliver high performance on these architectures, but they also need to work well when built and used in combination within the application. The extreme-scale scientific software development kit (xSDK) provides infrastructure for and interoperability of a collection of more than twenty related and complementary numerical libraries to support rapid and efficient development of high-quality applications.

This presentation will summarize the elements that are needed to make the xSDK an effective ecosystem of interoperable math libraries that can be built on top of large application codes. We will also discuss efforts to provide performance portability and sustainability, including xSDK testing strategies.

Invited Talks / 18

Fostering interdisciplinary research by composable Julia software

Author: Michael Herbst^{None}

Corresponding Author: herbst@acom.rwth-aachen.de

Today's ubiquitous data-driven workflows allow scientists to expand the limits of length and time scales in simulations. In my own field, namely first-principles atomistic simulations, the data itself is generated by systematic high-throughput workflows, which occupy a noteworthy chunk of the world's supercomputing resources. Questions related to the efficiency, robustness and accuracy of simulation protocols and the reproducibility of obtained simulation data are thus more pressing than ever. Due to the complexity of underlying physical models (non-linear PDEs, multi-linear algebra) tackling these issues is inherently an interdisciplinary endeavour. However, close collaboration of application scientists with researchers from mathematics or computer science requires software, which can support research thrusts all the way from model problems to full-scale applications.

I will discuss the opportunities of Julia programming language to satisfy the needs of interdisciplinary research. As an example I will focus on the density-functional toolkit (DFTK, <https://dftk.org>), a first-principle simulation code we started about 3 years ago. Being written entirely in Julia the code is highly accessible (only 7000 lines of code). At the same time Julia's composable programming paradigm allows (a) a seamless integration with standard HPC libraries (MPI, CUDA) and (b) to take advantage of unique features such as algorithmic differentiation. This has already enabled cross-disciplinary advances on error estimation or the developments of more robust algorithms. Notably, a number of these works involved undergraduates or PhD students from mathematics and computer science directly testing their work on relevant application simulations. Moreover the simplicity by which Julia enables code composability has stimulated joint initiatives to design common interfaces for sharing data within the young ecosystem. At the same time these efforts allow integrating with existing libraries outside Julia to avoid reinventing the wheel wherever possible.

Contributed Talks / 19

Graph-based Data Representation for Crash-worthiness Simulations

Authors: Anahita Pakiman¹; Jochen Garcke²

¹ *Fraunhofer SCAI-University of Wuppertal*

² *Fraunhofer SCAI, University of Bonn*

Corresponding Author: anahita.pakiman@scai.fraunhofer.de

We consider graph modeling for a knowledge graph for vehicle development, with a focus on crash safety. An organized schema that incorporates information from various structured and unstructured data sources is provided, which includes relevant concepts within the domain. In particular, we propose semantics for crash computer aided engineering (CAE) data, which enables searchability, filtering, recommendation, and prediction for crash CAE data during the development process. This graph modeling as an example for the overall CAE process considers the CAE data in the context of the R&D development process and vehicle safety. Consequently, we connect CAE data to the protocols that are used to assess vehicle safety performances. The R&D process includes CAD engineering and safety attributes, with a focus on multidisciplinary problem-solving. We describe previous efforts in graph modeling in comparison to our proposal, discuss its strengths and limitations, and identify areas for future work.

Today, the Finite Element method (FEM) is the preponderant tool for automotive crash simulation ¹. The large amount of complex data confronts engineers with the challenge to explore the simulation results sufficiently, due to lack of engineering time and limitations of data storage, processing and analysis tools. This need pushed the automotive companies to uptake preand post-processing tools to be more efficient in analysing the data, with the goal to spend the engineers time on solving the problem instead of data processing. Nevertheless, even with all achievements so far, data flow within the companies is still inefficient. Yet, crash scenarios studied in the development phase are just a tiny proportion of the real crashes. The need to increase the number of simulations and the limitation of CAE engineers' time emphasizes the importance of an intelligent system to capture domain knowledge as knowledge graphs (KGs) for automotive, which we call car-graph.

The modeling of CAE data is challenging since the data is complex, and several disciplines with different requirements interact with the CAE data. However, the flexibility of graph data modeling reflects existing uncertainties and allows the modeling to evolve. In this work, we present an initial attempt to define a semantic representation that stores information regarding the different crash scenarios, the vehicle design deviations during the development process, and the quantities of interest that measure the outcome. Consequently, we propose semantic selections that follow the development concepts, FE-modeling terminology, crashworthiness assessment quantities, and other relevant entities. Additionally, these can be used as input for machine learning (ML) analysis, where the graph modelling also allows storing ML results. Our vision is to use data modeling and ML to auto-assess the cause and effect in the development process to assist engineers and, in particular, to assess the safety of different, uncalculated crash scenarios.

As an example, we will present a summary of an industrial implementation for pedestrian analysis. Here, the number of simulations increases enormously for each design in pedestrian analysis. We will illustrate how CAE-web visualizes this big data and allows its intuitive and easy exploration. In this visualization, we present the traditional CAE reporting as a dynamic web interface and graph-ML technics on this data. We propose two groups of visualization: zoom-out and zoom-in views. Zoom-out views consider the assessment of many simulations, for example, development trees, status tables of safety performance, or embedded results from machine learning. However, zoom-in views contain single/multiple simulation assessments and comparisons. Additionally, the user has a multi-view functionality to combine zoom-in and zoom-out views. In multi-view, zoom-out views are selection inputs for updating zoom-in views.

Keywords: crash-worthiness; CAE data management; CAE knowledge; Car knowledge graph; data representation

References

1 P. Spethmann, C. Herstatt, and S. H. Thomke, "Crash simulation evolution and its impact on R&D in the automotive applications," *International Journal of Product Development*, vol. 8, no. 3, pp. 291–305, 2009.

20

Registration and Welcome

24

Reception

Workgroups / 28

Presentation of Workgroups

30

Conference Dinner

34

Lunch & End of Workshop

Invited Talks / 35

Creating sustainable research software by the example of the deal.II library

Author: Martin Kronbichler¹¹ *University of Augsburg***Corresponding Author:** martin.kronbichler@uni-a.de

In my talk, I will present the library deal.II, an open-source software aiming at the rapid development of simulation codes for partial differential equations based on the finite element method. The guiding principle of deal.II is to provide functions for the main building blocks in a solver that a user code can then combine and extend in an application-specific way. I will then give insight into my experience from starting or guiding several application projects that build on top of these abstractions. Across diverse scientific fields, spanning from geoscience over computational fluid dynamics to material sciences, there is a common mathematical underpinning that allows to re-use software concepts and contribute with new knowledge. A particularly important contribution of my work has been on the high-performance computing aspects of these projects, enabling solvers to run efficiently on current and evolving hardware architectures. We identified many possibilities to share knowledge and create synergies on the application side and to perpetuate our efforts by benchmarks as well as interaction with other big finite element projects.

Invited Talks / 36

FitBenchmarking: an open source tool for comparing data analysis software

Authors: Tyrone Rees¹; Andrew Lister²; Anders Markvardsen³; Jaroslav Fowkes²; Tim Snow⁴¹ *UKRI-STFC*² *STFC Rutherford Appleton Laboratory*³ *STFC Rutherford Laboratory*⁴ *Diamond Light Source***Corresponding Author:** tyrone.rees@stfc.ac.uk

STFC's Computational Mathematics Group provides support and mathematical software for the UK's large scale facilities, such as the ISIS Neutron and Muon source, the Diamond Light Source, the Central Laser Facility, and the Culham Centre for Fusion Energy. These facilities are visited by thousands of researchers each year, and they produce increasingly large amounts of data that needs to be processed. Furthermore, as the scale of data increases, it is more likely to need to be analysed without human intervention. Therefore it is more important than ever that scientists use the most robust and most efficient numerical algorithms.

Much of the data analysis that is carried out takes the form of fitting parameters to models, usually by formulating the problem as a nonlinear least-squares problem. Recently we have developed RALFit, a tensor-Newton nonlinear least-squares solver, and GOFit, a global nonlinear least-squares

algorithm. Alongside these we have developed FitBenchmarking: an open source python package which interfaces scientific data analysis software with a range of fitting back ends.

FitBenchmarking has been designed to help:

- Scientists, who want to know the best algorithm for fitting their data to a given model using specific hardware.
- Scientific software developers, who want to identify the best fitting algorithms and implementations. This allows them to recommend a default solver, to see if it is worth adding a new minimizer, and to test their implementation.
- Mathematicians and numerical software developers, who want to understand the types of problems on which current algorithms do not perform well, and to have a route to expose newly developed methods to users.

Representatives of each of these communities are involved in the design and implementation of FitBenchmarking.

The FitBenchmarking project embodies the FAIR principles, not only in terms of curated datasets we supply from a range of applications from across the UK's National Facilities, but also in terms of assisting scientists in finding cutting edge algorithms (and new implementations of algorithms). The tool has helped to foster fruitful interactions and collaborations across the disciplines and we plan to grow its reach further in the coming years.

Contributed Talks / 37

Building a Knowledge Graph for Scientific Computing

Authors: Frank Wübbeling^{None}; René Fritze^{None}

Corresponding Author: frank.wuebbeling@exchange.wwu.de

Mathematical computing knowledge is produced at an immense, and seemingly ever increasing, speed. Very little of it is organised in meaningful ways, making its discovery, insight and discussion harder every year. Following new developments in a given field is time consuming even for experts. Entering a new specialisation is daunting for students.

We will show how building a knowledge graph for scientific computing can address these issues. We have created an ontology that semantically links mathematical problems with publications, algorithms and implementations.

The ontology encodes possible relationships between the entities in the graph. These connections can be explored using a web-based query frontend, which enables non-experts to quickly gain an overview of available methods and software for specific numerical problems in their scientific work. For experienced users, it makes variations of existing algorithms easily discoverable and allows tracking of new publications or software implementations connected to a specific problem.

We will be inviting feedback for our plans to grow this knowledge graph into a community-driven platform with an open, freely accessible API.

Our efforts are part of the scientific computing task area in the Mathematical Research Data Initiative (MaRDI), a consortium in the German National Research Data Infrastructure (NFDI).

Contributed Talks / 38

Design of a Workflow Description for Documentation and Integration of FAIR Computational Experiments

Authors: Pavan Veluvali¹; Jan Heiland¹; Peter Benner¹

¹ *Max Planck Institute for Dynamics of Complex Technical Systems*

Corresponding Author: veluvali@mpi-magdeburg.mpg.de

Numerical algorithms and computational tools are essential for managing and analyzing complex data processing tasks. With increasing meta-data awareness and parameter driven simulations, the demand for reliable and automated workflows to reproduce computational experiments across platforms has grown.

In general, computational workflows describe the complex multi-step methods that are used for data collection, data preparation, predictive modeling, and simulation in various engineering applications. They are characterized through their input-output relation such that the associated meta-data can be used interchangeably and redundantly.

In this regard, we develop a prototypical CSE workflow that abstracts the multi-layered components from computational experiments. As a case study, we incorporate the time-dependent Stokes-Darcy solver 1 into our workflow, and execute the coupled system of free flow adjacent to a permeable porous media via the monolithic block-preconditioning scheme implemented entirely in the DuMux framework. Within this example, we focus on solver approaches for the coupled problem, and determine the run time and memory behavior of the system. Moreover, the workflow adheres to FAIR principles, such that abstracted components are Findable, Accessible, Interoperable, and Reusable 2. Lastly, we discuss how the CSE workflow description presented here as a part of the MaRDI consortium serves as a scientific tool for research data management in numerical mathematics.

References:

1 Schmalfuss, J., Reithmueller C., Altenbernd M., Weishaupt K., Goeddke D., “Partitioned Coupling vs. Monolithic Block-Preconditioning Approaches for Solving Stokes-Darcy Systems.” *arXiv preprint arXiv:2108.13229* (2021).

2 Carole G., Sarah C-B., Stian S-R., Daniel G., Yolanda G., Michael R. C., Kristian P., Daniel S., FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121.

Invited Talks / 39

ADOL-C: 40 years of software development

Author: Andrea Walther^{None}

Corresponding Author: andrea.walther@math.hu-berlin.de

The provision of derivatives for a function defined by an evaluation procedure in a high level computer language like Fortran or C forms an important task for numerous applications comprising for example optimization, parameter estimation, and data assimilation. The technique of algorithmic differentiation (AD) offers an opportunity to provide derivative information of any order for the given code segment by applying the chain rule systematically to statements of computer programs.

The package ADOL-C uses operator overloading for differentiating automatically C and C++ codes. During an evaluation of the function to be differentiated the usage of a new data-type `adouble` causes the generation of an internal function representation. Afterwards several drivers allow a very flexible choice of the mode and order of differentiation to be performed. Naturally, this approach works also for codes based on classes, templates and other C++-features. The resulting derivative evaluation routines may be called from C, C++, Fortran, or any other language that can be linked with C.

In this presentation we briefly present these features of ADOL-C together with important applications of ADOL-C. This will go along with an extensive overview of 40 years of software development in various research environments discussing some of the challenges that ADOL-C faced during this period.

Contributed Talks / 40**Towards a Benchmark Framework for Model Order Reduction in the Mathematical Research Data Initiative (MaRDI)****Authors:** Peter Benner¹; Christian Himpe^{None}; Kathryn Lund^{None}; Jens Saak¹¹ *Max Planck Institute for Dynamics of Complex Technical Systems***Corresponding Author:** lund@mpi-magdeburg.mpg.de

The race for the most efficient, accurate, and universal algorithm in scientific computing drives innovation. However, this healthy competition is only beneficial if research outputs from different projects are actually comparable to one another. Fairly comparing algorithms can be a complex endeavor, as the implementation, configuration, compute environment, and test problems need to be well defined. Due to the increase in computer-based experiments, new infrastructure for facilitating the exchange and comparison of new algorithms is also needed. To this end, we propose a benchmark framework, which is a generic toolkit for comparing implementations of algorithms using test problems native to a community. Its value lies in its ability to fairly compare and validate existing methods for new applications, as well as compare newly developed methods with existing ones.

As a prototype for a more general framework, we have begun building a benchmark tool for the Model Order Reduction Wiki (MORWiki). The wiki features three main categories: benchmarks, algorithms, and software. An editorial board curates submissions and edits entries. Data sets for linear and parametric-linear models are already well represented in the existing collection. Data sets for non-linear or procedural models are being added and extended. Searchable attributes for all categories are actively being aggregated in metadata databases.

The MORWiki collection will serve as the primary basis for our model reduction benchmark tool. To this end, experiences from related projects serve as prototypes and will be extended to encompass diverse model types and performance measures. The MORWiki will serve as a proof-of-concept for a living document and progress-tracker of a field, while also facilitating fair comparisons of new findings and methods. Its core information will be mirrored in the MaRDI-Portal, which is concurrently under development.

Invited Talks / 41**MaRDI - The Mathematical Research Data Initiative within the German National Research Data Infrastructure (NFDI)****Author:** Thomas Koprucki¹**Co-author:** Karsten Tabelow¹¹ *Weierstrass Institute for Applied Analysis and Stochastics***Corresponding Author:** thomas.koprucki@wias-berlin.de

Like in all scientific disciplines research data in mathematics has become vast, it is complex and multifaceted, and, through the successful application of mathematics in interdisciplinary research, it is widespread in the scientific landscape. It ranges from information bases such as the standard reference data for special functions, tables and similar mathematical objects to highly complex data in scientific computing or scientific machine learning. The growing amount of research data challenges an old requirement in science: its reproducibility and the re-usability of results. In an attempt to answer this challenge at current level, the FAIR principles have been formulated. Yet, despite the existence of special solutions a comprehensive infrastructure for research data in science or in mathematics is missing that supports the research process and implement the FAIR principles. Thus, the German Council for Scientific Information Infrastructures initiated the foundation of the German National

Research Data Initiative (NFDI) to address the need for discipline specific research data infrastructures and to conform to the specifications of the European Open Science Cloud (EOSC).

In this context the Mathematical Research Data Initiative (MaRDI) within the NFDI aims at developing a research data infrastructure for mathematics. Starting with the areas of computer algebra, scientific computing, statistics and machine learning, MaRDI will develop standards for confirmable workflows and certifiable mathematical results and provide new services that assist the research cycle up to peer-review in the publication process. Standardised formats, data interoperability and application programming interfaces need to be established to ensure the ease of use of data across broad disciplines. Furthermore, by building the MaRDI portal as a decentralized and federated infrastructure the storage and accessing of data and knowledge will be facilitated in a manner that would perpetuate FAIR and open science principles.

In this talk, we will give an introduction into the notion of mathematical research data and the use cases for a corresponding infrastructure in the mathematical research process and the emerging national and international research data landscape. We present MaRDI's concepts and ideas to implement the FAIR principles for mathematical research data and illustrate them with examples.

Invited Talks / 42

Metadata4Ing: An ontology for describing the generation and provenance of research data within a scientific activity

Authors: Dorothea Iglezakis^{None}; Benjamin Farnbacher¹

¹ *TU München*

Corresponding Author: benjamin.farnbacher@tum.de

Knowledge graphs basing on ontologies enables us to describe and connect research data, software, methods, actors and instruments in a machine readable and actionable manner. Ontologies function in this context as a formalized language that unify a semantic description of research results, their content and their provenance.

Metadata4Ing (m4i) (<https://w3id.org/nfdi4ing/metadata4ing/>), developed within NFDI4Ing, is a mid-level ontology that provides a framework for the semantic description of research data, with a particular focus on engineering and neighbouring disciplines. It offers terms and properties for the description of scientific workflows and research results. It considers, for example, the object of investigation, sample and data manipulation procedures, a summary of the data files, and personal and institutional roles of participants in data-driven research processes. The terms of Metadata4Ing are available on the terminology service of NFDI4Ing (<https://terminology.nfdi4ing.de/ts4ing/ontologies/m4i>) and on Linked Open Vocabularies (<https://lov.linkeddata.es/dataset/lov/vocabs/m4i>).

Metadata4Ing builds on existing ontologies like the Basic Formal Ontology (BFO), the PROV Ontology and the Data Catalogue Vocabulary (DCAT) and is extendable to the requirements of specific fields by deriving subclasses with specific properties. A subontology for high performance computing (HPC) workflows is currently in development.

The description of engineering processes can benefit greatly from the joint usage of m4i together with the existing and planned databases and knowledge graphs for models, algorithms and software developed within MaRDI. This talk will present the basic concepts and data model of m4i, its connection points to MaRDI and the prospective usage of m4i to describe HPC processes and results.

Invited Talks / 43

preCICE – A General-Purpose Simulation Coupling Interface

Author: Benjamin Uekermann¹

¹ *University of Stuttgart*

Corresponding Author: benjamin.uekermann@ipvs.uni-stuttgart.de

preCICE is an open-source coupling software for partitioned multi-physics and multi-scale simulations. Thanks to the software's library approach (the simulations call the coupling) and its high-level API, only minimally-invasive changes are required to prepare an existing (legacy) simulation software for coupling. Moreover, ready-to-use adapters for many popular simulation software packages are available, e.g. for OpenFOAM, SU2, CalculiX, FEniCS, and deal.II. For the actual coupling, preCICE offers methods for fixed-point acceleration (quasi-Newton acceleration), fully parallel communication (MPI or TCP/IP), data mapping (radial-basis function interpolation), and time interpolation (waveform relaxation). Today, although being an academic software project at heart, preCICE is used by more than 100 research groups in both academia and industry. The wide variety of application fields ranges from aerodynamics to astronautics, automotive manufacturing, wind energy, biomechanics, biomimetics, marine engineering, nuclear fusion, reactor safety, geophysical systems, and many more.

Contributed Talks / 49

Towards Foundations for Open Interfaces for Scientific Computing

Author: René Fritze^{None}

Co-authors: Mario Ohlberger ; Stephan Rave

Corresponding Author: rene.fritze@exchange.wwu.de

Algorithms and models realized by established software packages can be hard to exchange, compose or interconnect in the context of complex modeling or simulation workflows.

In this contribution we will present our work towards developing and establishing open interface standards with a core API toolkit.

These open interfaces will improve the reusability of numerical models and facilitate their recombination in complex simulation workflows. By enabling researchers to reuse existing realizations of numerical models, significant development time can be saved, while collaboration between experts in different fields of scientific computing is fostered. Moreover, interface standards for numerical models improve the comparability of numerical methods by facilitating computations with competing algorithms for the same model.

We will showcase a prototype C-language based API toolkit that allows accessing interfaces implemented with the toolkit by loading the other software component as a shared library plugin. Language bindings for this toolkit are implemented for C, C++, Julia and Python.

Our efforts are part of the scientific computing task area in the Mathematical Research Data Initiative (MaRDI), a consortium in the German National Research Data Infrastructure (NFDI).